CENTRALIZED COMPUTER RESOURCES IN EXPERIMENTS ON HIGH ENERGY PHYSICS

A.E. Shevel [1], A.N. Lodkin [2], A.A. Oreshkin [3]

Introduction

Here we would like to cover briefly the key points of centralized computer infrastructure in High Energy Physics (HEP) experiments of PNPI RAS. The features of providing HEP experiments by computer resources were repeatedly considered (e.g. R.M. Brown et al., IEETF, 1966. P. 114; H.E. Devies, 1976 CERN School of Computing, P.161). In publications devoted to computer aspects of scientific research the automation of specific technical tasks and the methods of their solution are often discussed. We would like now to point out important stages of achieving a common goal which was pursued by all papers discussed in this review, namely: to increase the efficiency of man-machine systems in scientific information processing in the case of HEP experiments at PNPI.

We will bear in mind that increasing efficiency includes a lot of components. In other words, it is meant increasing volume of really available possibilities for scientific information processing at the same or less price, which can be expressed in different ways:

- in increasing technical and functional capabilities of man-machine system as a whole;
- in increasing reliability of hardware and software;
- in reducing operating staff necessary for supporting computer complex in operational state.

It is supposed that while improving one parameter of the system other parameters do not change or change in the direction of increasing efficiency.

The source of scientific data in a nuclear physics experiment is a measuring experimental setup. For the data to be processed on centralized computer capacities, they can be acquired through communication lines or read from secondary memory where they were stored earlier.

The use of centralized computer resources is caused by a relatively high cost of a computer complex and a large number of users, which varied from 10 to 200 people at different times. There is no doubt that centralized computer complex with unique technical parameters within the institute (or division) framework always existed, exist and will exist. Only the way of integration of such complex into the institute (division) computer infrastructure may be a subject of discussion or optimization.

The centralized computer complex is designed not only for computing, but also for effective using more expensive resources than the complex itself: the PNPI synchrocyclotron operation and working time of PNPI researchers. The latter determined an important feature – the centralized computer was used in two modes: for serving experiments on-line and for typical running batch jobs. On-line mode was always of higher priority than running batch jobs.

The use of a computer for on-line work supposed that a program complex was run, which was used for specific physical experiment. Typically program tasks were:

[1] e-mail:*shevel@pnpi.spb.ru*
[2] e-mail:*lodkinan@pnpi.spb.ru*
[3] e-mail:*oreshkin@pnpi.spb.ru*

- acquiring information from the experimental setup and storing it on a mass storage (as a rule, on magnetic tapes);
- partial or complete processing of information acquired;
- displaying partial results.

Here it is important to note the volume and the intensity of experimental data acquiring. Of course, these values varied depending on the experiment character, however in many cases the volumes were estimated at hundreds megabytes per day. Typical measurements duration was from 1–2 days (debugging and calibration) to 20 days and more.

Below the main basic technological problems (which were due to the available generation of computer technique) and methods of their overcoming are discussed.

First computers of the third generation

In autumn of 1973, the computer ES–1020 with the main memory of 64 kbytes was put into operation. It had two disk drives with 7.25 Mbytes each, four tape bobbin drives (24 Mbytes on bobbin), a card reader, a card punch, a line printer. The average capacity on mixture of instructions gave about 20 thousands instructions per second. The Disk Operating System for computers of Edinoi Serii (DOS ES) was run; it had a number of compilers for such languages as Fortran, PL/I, COBOL, and other typical program components: all sorts of utilities, linkage editor (task builder) and so on.

Note the problems facing at that time.

- DOS ES version used in period 1973–1974 was poorly suitable for acquiring data from experimental setups. In particular there were limited capabilities of processing non-standard devices interruption.
- The limited space on disks and in a main memory also made it much difficult to transfer applied programs which were available, for example, at CERN.
- The serious problem was the insufficient reliability of hardware (tape drives, disk drives). Naturally, it caused the fear to lose software or data.

To a considerable extent the same problems remained also after putting into operation the next computer complex consisting of two computers ES–1030 in the middle of 1977. It had 16 disk drives with 7.25 Mbytes each, 8 tape drives and other peripherals. It took almost 200 m^2 of premises for deployment of this hardware, and the labour of 20 people (hardware engineers, programmers, operators) was required to maintain this complex in operational state. The computing complex on the base of ES-1030 computers was employed as a centralized computing resource in experiments till 1989.

At first stages of using computers of the third generation in physics experiments, it was necessary to create a software-hardware base for using the computers of such class in performing measurements. So, the special telecommunication I/O controller for transmitting data from experimental setup to central computer was developed, which enabled to connect up to 16 experimental setups [1]. By this step a hardware base was created for a reliable link between

ES computer and mini-computers (SM–3, SM–4) which were operational elements of all experimental setups. Architecturally this solution allowed to have relatively simple mini-computers be included in experimental setups. On the other hand, relatively expensive and complicated in servicing devices, such as high speed bobbin tape drives and a high capacity disk storage (of that time), could be concentrated on the centralized computer.

In order to make computer infrastructure more reliable and friendly to man, the developments were undertaken in the area of system software and hardware architecture.

One of the first works on software for computers of such a class was the supervisor modernization (a resident part of the operating system) [2].

For increasing reliability of work with tape drives, a special algorithm was developed, which prevented a loss of experimental data in most cases of device hardware failures and operator errors: incorrect tape positioning, invalid tape bobbin [3].

For improving convenience of working with secondary storage on disks, a special access method was implemented, which made it possible to consider secondary storage as a main memory with random access [4]. The data exchange was performed by calling appropriate subroutines with parameters.

The necessity to improve the quality of interface between man and computer led to the development of the dialog system [5] which was the basic one during a number of years. In this system a dialog initiated by computer (mode "menu") and a set of patterns for preparing user programs of serving a specific experiment were used. The system enabled to implement quite complicated scenarios of work in an experiment, it provided a maximum usage of all centralized computer resources available at that time.

The great significance was attached to the development of applied program system which represented sets of programs optimized on resources used and degree of stability to hardware failures. These systems implemented complete simple scenarios which were easily formulated in substantial terms, for example:

1. Acquiring a portion of experimental data from a communication line.
2. Writing a data portion acquired on a magnetic tape.
3. Writing a data portion acquired on a disk.
4. Providing synchronization with an user data processing program.
5. Go to the point 1.

Here it is supposed that the user program having a possibility to read experimental data from a disk and to process them may be (or not) started. The synchronization is needed to prevent processing of the same data more than once.

For implementing such scenarios several applied systems were developed: providing various in volume service for experimentalist, they consumed at the same time different shares of centralized resources according to provided opportunities [6]. Thus, there was a quite large number of programs which could be used in a specific experiment. One example of using centralized computer infrastructure is given in Ref. [7].

In the period 1980–81, it became obvious that the bottle-neck of centralized computer resources was preparing and debugging programs. Most of programs was punched and introduced into computer using card reader. On the other hand, in experiments the electronic means of displaying were widely used, which were connected mainly to mini-computers of SM class. All these circumstances stimulated the development of a local terminal network [8]. A computer SM–4 connected with several tens of terminals was used as a controller of a local terminal network. The terminals could be connected at the distance of up to several hundreds meters of two twisted pairs, which made it possible to place terminals where necessary. The important feature of terminal system was that any terminal could be logically connected to any computer: ES–1030/1, ES–1030/2 and SM–4. This enabled to prepare and debug programs for several computers literally without rising from table.

The experience of program development has led naturally to employing the special technology of developing large programs and program systems [9], which was used in a number of works [5, 6, 8]. The technology named metaprogramming allows to describe an architecture of large program systems regardless of programming languages used and computer type. In metaprograms only minimum set of operators is used: check, conditional transfer, transfer of control, comment, subroutine call. Thus quite simple means enable to describe a logic of complex program systems.

Among others, there was an important problem of long-term storing program texts, descriptions and other documents on computer media [10]. It is a matter of archive scenario in a traditional sense: to write the same programs or data for long-term storing (many days, months or years) when necessary.

Main capabilities of this program package are as follows:

- writing information on a magnetic archive tape in OS ES standard with possibility to append information to a partially filled archive tape;
- reading any module or group of modules by name, date, version number and so on from an archive tape and writing on another tape or disk;
- at last, this system had special methods of error processing, which enabled to restore information even in the case of very severe magnetic tape failures.

Virtual Machines System

In 1988 the first of two ES-1046 computers was installed. It used the more powerful operating system – the Virtual Machines System (VM).

The network of terminals (about 70) was developed. They were mainly alphanumeric terminals of type ES–7970 and ES–7920. Although computing resources were insufficient for simulating complex experiments, the terminal network gave possibility of simultaneous work for many scientific workers and engineers (to prepare papers, descriptions, as well as to formulate problem solutions on the language of computer programs).

In the second half of 80s there was no doubt of future wide using of personal computers (PC). In order to simplify transition from traditional computers to PC there were used PC running emulation programs which enabled to emulate terminals ES–7970, to copy files from ES–1046 to PC and back, as well as to use ES–1046 disk storage as an additional PC disk drive. Thus the possibility of smooth transition to desktop computers was provided.

One of main problems, which we were faced with in conditions of considerable growth of users number, was a low reliability of data storing on disk media. It was caused by unreliable hardware and software components of a computer setup. As any information stored in the secondary memory (programs or data) represents by itself materialized working time, therefore a loss of information is an obvious large waste for the institute.

With the goal of decreasing the probability of information loss, we have undertaken the development of program complex [11] which would enable to make backup copying (including incremental backup copying) and restoring this backup program complex under full control. Making backup copying or restoring did not require a skilled worker. Any worker who was able to mount tape in tape drive could perform any operations: the backup program itself said which tapes should be installed and controlled the correctness of tape mounting. During execution of one or another operation, the protocol was logged using which not only all actions of operator but the status of individual directories could be restored. Such mechanism was used from 1989 till 1992 on ES–1046 and later – on communication and computing servers.

Thus, the quite comprehensive backup system was developed; in such routine operations as copying and restoring this system left for the staff only absolutely necessary minimum, for example the choice of a tape drive.

Operating system VM gave a lot of possibilities to the operating staff: it was possible in a number of cases to run different operating systems at the same time for different users, to test equipment while system operating and so on. However, for users no qualitative changes happened in character of using centralized computers. A computer was mainly used as a big calculator.

UNIX

The transition to new operating platforms (systems like UNIX) for workstations of high and middle performance was predetermined by appearance of microprocessors manufactured on RISC technology. Since many different types of microprocessors (MIPS, DEC, HP and others) became available, to minimize expenses on basic software the operating platform had to be relatively easily installed on machines with different architecture preserving at the same time all wealth of capabilities. By that time the system UNIX was the operating platform which met such requirements. Up to now the operating platform UNIX is serving de facto as a standard, i.e. it is the universal program platform for all workstations and many supercomputer installations.

The gradual transition to new platforms (hardware and software) took place around the world in all areas of using computers including scientific area. So, at CERN traditional architecture computers mainframes (such as ES series computers) were closed down in 1995, at DESY – in 1996. At PNPI this transition took place with a rapid rate due to a lack of financial resources for a gradual transition to new platforms and finished in the late 1992.

For the first time UNIX (Xenix) became to be available for users of PNPI in 1989 on PC/386. CERN library was installed on this computer. Then the programs installed on PC/386 were transferred to PC/486/33 [12], and now they are used on the centralized communication server.

The important point in transition from centralized systems of mainframe type to computers based on microprocessors was that this transition was made by the same specialists who had already the experience of serving a large number of users on centralized computer setups that in turn saved time and labour of experimentalists during period of transition to new computer platforms.

Local network

In the very early stages of using computers in nuclear physics, it was assumed that centralized computer complex had to be connected to experimental setups. With the purpose of connecting computers into united coherent environment, experiments in the area of local networks were also carried out. The first Ethernet network was developed in 1990 ("MIFIR" network). The network consisted of 4 nodes: ES–1046, SM–1300, SM–4, DVK–1. Almost at the same time the new comprehension of the institute computer infrastructure was shaped [13].

The first Ethernet TCP/IP network segment based on thin coaxial 50 Ohm cable was realized in 1993. There were three workstations and several X-terminals in the segment. At present, the institute scientific local network is based on Ethernet 10 Mbit/s technology with the UTP cabling system. The interfaces between main institute buildings and the central network node are implemented using fibre optic cables. There are more than 300 nodes in the institute network. The main network protocol is TCP/IP.

Centralized servers

The PNPI computer infrastructure is moving towards specialization of centralized servers connected to the institute local network. Below we enumerate briefly available servers.

Communication server of the institute is implemented on the base of Intel/Express Pentium/60, memory 64 Mbytes, EISA, modem battery, disks 5.3 Gbytes, Exabyte 8200, floppy 3.5", CD ROM. About 340 users are registered on the server. About 200 people use the server every month.

Computing server is implemented on the base of SGI (R3000), memory 224 Mbytes, three SCSI adapters, disks 16.4 Gbytes, Exabyte 8500, 6 processors. 80 users are registered on the server and more than 60 users use the server every month.

Data base server is implemented on DEC Alpha 4/1000 266 MHz, memory 128 Mbytes, disks 14.9 Gbytes, CD ROM. The server was installed in the early June of 1996. The server is supposed to be used by all institute laboratories for storing the institute scientific data base.

Internet

For a long time experimental works in the area of a computer communication with large foreign physics centres were carried out at PNPI. So, in 1987 more than ten communication sessions with CERN were carried out. The communication was performed from the alphanumeric terminal via modem 2400 bit/s to Petersburg (SPIIRAN), then through network AKADEMSET to Moscow (node VNIIPAS), further via dedicated telephone line to communication node in Vienne (Austria). The arrangement of every session required most qualified attention and as a whole appeared to be a nontrivial procedure.

At last, in March of 1994 the first IP channel on the base of switched telephone line "ISKRA–2" [PNPI–ITEP, later Radio–MSU (Moscow)] was realized; it gave possibility of using the world computer network Internet. The real achieved capacity of line "ISKRA–2" [Gatchina–Moscow] with using modems Telebit World Blazer was 12 kbit/s.

Since 23 September of 1995 PNPI has the satellite IP channel with bandwidth 64 kbit/s (full duplex), which was implemented with the help of international scientific foundation INTAS according to the project INTAS 93-56. It enabled to use widely new possibilities having opened for the institute in Internet.

The main institute task in an information area is a conversion of information about the matter structure from natural form into human one, i.e. articles, books, manuals and so on. It is natural that effective ways of processing data of all kinds using computer are a necessary component of scientific research. However, making decision on processing scientific information is performed by human, but not by machine. One of the powerful stimulus for more quick making decision is an information exchange between researchers. Such exchange is fulfilled according to various scenarios which are realized by appropriate software: mail, ftp, news, www, irc, ytalk, white board, vat (audio), nv (video) and so on. [14].

Main directions

As it is seen from a brief review of works made on developing centralized computer resources, a computer infrastructure never can be completely accomplished. In connection with this it is useful to note perspective areas in PNPI computer infrastructure.

- It is natural that for the convenience of processing all information has to be represented in an electronic computerized form. As there is a lot of electronic forms, the tools and methods for free conversion from one electronic form into another should be provided. There is no doubt that one should have a wide range of equipment for conversion from traditional forms of representation into electronic and back – scanners, printers, video and audio recording and playing back apparatus. Among them there is, of course, an equipment for seminars and presentations.

- It is important to store data archives on well protected and preserving media with a convenient network access to it.

- The Internet channel of a sufficient bandwidth is a necessary condition of participating of the institute in international scientific labour sharing.

  According to an estimation, the institute should have up to 2000 year a main Internet channel with a bandwidth no less than 2 Mbit/s (full duplex).

- The information infrastructure (infostructure) of the institute – Intranet.

  The availability of appropriate software for supporting a variety of data exchange scenarios makes it possible to use these tools for internal institute needs, such as institute announcement board:
  – seminars;
  – scientific and coordinating boards: announcements and decisions;
  – administrative instructions, orders;
  – technical information: schedule of work of reactor, accelerator, warehouses, publishing and foreign departments, other service;
  – telephone book;
  – private announcements, other information.

- Virtual reality.

  The virtual reality tools enable to shape on a workstation screen a three-dimensional picture of being built or already built object from any point of view (experimental device, building, nuclear reactor body and etc). Changing a point of view it is possible to "travel" inside an object and to view the interior which is formed dynamically on the workstation screen according to working drawing with the speed of changing view point. It permits the experts to travel through tubes, crossing, walls and so on.

Conclusion

The qualitative differences between existing centralized computer resources and the mainframe stage, which came to the end at PNPI in 1993, make no doubts for anybody. It is useful to note what they show themselves in.

As early as in the 80s it became clear that for building of computer infrastructure one should use ready products (software and hardware) which present on a computer market or free available, such as GNU packet or CERNLIB. In other words, it is not perspective to try to make own developments of small components and to adjust them for some special features of computing environment; the ready products should be used as they are. New developments and projects in the area of centralized computer resources should be made primarily in choosing available solutions of lower level (workstations, servers, routers, network standards, programming systems, robotized peripherials and so on), as well as in the sphere of infostructure.

A computer is a flexible tool for communicating between people and for representing data in convenient for perception form: text, graphic, audio, video and so on. Fax, autoanswerer, TV set, telephone, radio set, video tape recorder, device for audio playing back – all these devices are easily combined in a personal computer. Using a computer without a network is considered to be unreasonable, like transportation of a car using the horse power. However, while estimating the projects of changing the computer infrastructure integral expenses on realization (including the cost of a further maintenance) should be taken into account. One of the consequences of such statement is the conclusion that for the most of users it is more technological to use an X-terminal instead of a computer.

Since the volume of information in WAN (Wide Area Network) is more than in any library of the world, so one of important new infrastructure element is special tools for importing necessary information from WAN and exporting information to networks of different types and scales.

At last, the most important difference is a growth of the number of scenarios of interactions between people in which computer is used as an operational element.

References

[1] *B.V.Bochin, Yu.V.Elkin, P.V.Neustroev.* Preprint LNPI–284, Leningrad, 1976. 18p.

[2] *A.E.Shevel.* Preprint LNPI–128, Leningrad, 1974. 8p.

[3] *M.N.Andronenko, I.I.Gracheva, A.E.Shevel.* Preprint LNPI–658, Leningrad, 1981. 24p.

[4] *I.I.Gracheva, A.A.Oreshkin, A.E.Shevel, V.A.Schegelsky.* Preprint LNPI–496, Leningrad, 1979. 16p.

[5] *T.S.Serebrova, L.F.Suvorova, A.E.Shevel, S.V.Yazikova.* Preprint LNPI–302, Lenigrad, 1977. 34p.

[6] *A.N.Lodkin, B.Yu.Sokolovsky, A.E.Shevel.* Preprint LNPI–1063, Leningrad, 1985. 30p.

[7] *S.S.Volkov, Yu.A.Dotsenko, L.G.Kudin, N.P.Kuropatkin, A.A.Oreshkin, O.E.Prokofiev, A.E.Shevel.* Preprint LNPI–674, Leningrad, 1981. 42p.

[8] *A.G.Atamanchuk, V.I.Benevelsky, I.I.Gracheva, N.M.Gulina, S.G.Dolgobrodov, A.N.Lodkin, P.V.Neustroev, A.A.Oreshkin, E.M.Orischin, T.S.Serebrova, N.A.Seregin, B.Yu.Sokolovsky, E.V.Fotieva, A.E.Shevel.* // Avtometriya, 1986. No. 4. P.45.

[9] *A.E.Shevel.* // Programmirovanie, 1986. No.1. P.31.

[10] *I.I.Gracheva, A.N.Lodkin, A.E.Shevel.* Preprint LNPI–914, Leningrad, 1983. 21p.

[11] *S.A.Kulikov, A.E.Shevel.* // Upravl. Sistemy i Mashiny, 1991. No.6. P.133.

[12] *A.N.Lodkin, A.A.Oreshkin, T.S.Serebrova, A.E.Shevel.* // Proc. Int. Workshop on Programming and Mathem. Methods of Solving Physical Tasks, Dubna, 1993. P.114 (in Russian).

[13] *A.E.Shevel.* Preprint LNPI–1647, Leningrad, 1990. 12p.

[14] *A.E.Shevel.* // Proc. the Jubilee 30th PNPI Winter School, St. Petersburg, 1996. P.247 (in Russian).

2. NUCLEAR PHYSICS